

Speech Processing 2009/10

2nd Test

May 17th 2010

Please identify this form with your name and student number in the reserved space at the bottom. The answers to multiple-choice questions will only be accepted if inserted in the appropriate place. Wrong answers will be penalized. The phonetic symbols should use the SAMPA alphabet (Lisbon accent).

1. Classify as True (T) or False (F)
 - (a) Signal to noise ratio is a good objective measure for the evaluation of the quality of MBE coders.
 - (b) The logarithmic version of PCM explores the fact that the human ear is not so sensitive to noise, if the signal energy is small.
 - (c) The long-term predictor adopted in APC coders uses 2 or 3 coefficients as the pitch period does not always correspond to an integer number of samples.
 - (d) CELP coder versions standardized for 2.4kbps outperform MELP at the same bit rate.
 - (e) Search in the stochastic codebook of CELP coders is done using totally quantized predictors.
2.
 - (a) How many pulses does multi-pulse typically need per 10ms to produce good quality (without using long-term prediction)?
 - (b) What is the typical range of values of the parameter that controls the noise shaping filter in this family of coders?
3. Give examples (one per class) of
 - (a) a waveform coder that transmits side information
 - (b) a coder with bit rate lower than 800 bps
 - (c) a coder which models jitter in the synthesizer
 - (d) a coder using vector quantization and perceptual weighting
 - (e) a coder not using linear prediction to model the spectral envelope
4. The normalized version of RPE-LTP at 13kbps is based on frames of 5ms, for the excitation (6 bits for the normalized amplitude, 3 bits per pulse), and the long-term predictor (2 bits for the scale factor), and 4*5ms for the short-term predictor (36 bits per frame). Show the bit assignment for each type of parameter, and the total bit rate with frames of 7.5ms, maintaining the same density of pulses.
5. Classify as True (T) or False (F)
 - (a) Formant synthesizers use a two-mass model of the vocal cords, where the masses are coupled with a string.
 - (b) The harmonic component in H+N models is a sum of all the harmonics of the signal.
 - (c) Duration rules may be learned from data
 - (d) In PSOLA, a pitch modification also implies compensating the implicit duration modification.
 - (e) The greedy algorithms that CARTs use (for instance, to acquire intonational phrasing rules) are suboptimal.

- (f) Semantically unpredictable sentences are typically used to evaluate the performance of the prosody generation module.
6. Indicate one strong point and one weak point about formant synthesizers.
 7. Indicate 3 pairs of homographs for which the resolution of ambiguities in morphological classification is essential for the grapheme-to-phone conversion (you may do that for English if your native language is not Portuguese).
 8. Grapheme-to-phone approaches based on neural networks require a preliminary alignment of the training data. Indicate two graphemes which may be matched with a phonemic null in this alignment step, together with words where these nulls may occur. (You may do that for English if your native language is not Portuguese).
 9. The word to be synthesized is *caso*.
 - (a) The available corpus has the words *casar*, *fazer*, *acato*, *fase*, *faço*, *russo*, *uso*, and *mudo*. Indicate which phone sequences you will select from which words to concatenate the target word.
 - (b) The prosody generation module computed a target duration of 90ms for the fricative sound. The left sub-unit has a duration value of 80ms and the right sub-unit has a duration value of 40ms. Indicate the target durations for these two sub-units of the fricative.
 10. Write two pronunciations for the following foreign names and acronyms. The first pronunciation should be the one typically heard in Portuguese media. The second pronunciation is the one obtained using the grapheme-to-phone rules for the common lexicon in European Portuguese. Both should use only the phonetic SAMPA symbols of this language.
 - (a) generation
 - (b) mouse
 - (c) ADSE
 11. Using the following syntax

$$a \rightarrow b / c \dots d$$
 describe the simplified rules for grapheme *a* (unstressed) for Brazilian Portuguese, where *c* and *d* may be graphemes (e.g.: *a*, *b*, etc.), classes of graphemes (vowels, consonants, etc.), the word boundary (#), or any grapheme (*). You may use the symbol *0* to mark phonemic nulls, and the symbols | and () to mark disjunction between several graphemes (e.g.: *a | b | c*). The rules are applied in order, until one matches the context and, in this case, the following rules are not applied. The “simplified” rules do not need to contemplate compound words. In fact, they should only account for the cases depicted in the examples below: *comam*, *mandei*, *canhão*, *muda*, *mudas*.

Test 2 - Answers

Name:	
Number:	

1. (1.5 val.) Indicate T or F:

a	b	c	d	e

2. (1.2 val.)

a	
b	

3. (2.5 val.)

a	
b	
c	
d	
e	

4. (2.3 val.)

long-term predictor	
short-term predictor	
excitation	
total	

5. (1.8 val.) Indicate T or F:

a	b	c	d	e	f

6. (1.2 val.)

strong point	
weak point	

7 to 9 (1.2/1.2/2.0 val.)

7	
8	
9	

10. (2.1 val.)

transc	generation	mouse	ADSE
transc1			
transc2			

The answer to the last question (11) should be written in the next page (3 val.).